

## TABLE OF CONTENTS

2.2	Rainfall.....	39
2.2.1	Quality Assurance/Quality Control of Rainfall Data.....	39
2.2.2	Transformation to Grid-Based Data Set .....	43

## LIST OF FIGURES

Figure 2.2.1.1	Location of Rainfall Stations.....	40
Figure 2.2.2.1	Example of TIN-10 Estimation for Model Grid Cell .....	44
Figure 2.2.2.2	Grid Values of Annual Average Rainfall .....	45
Figure 2.2.2.3	Monthly Mean with 10 <sup>th</sup> and 90 <sup>th</sup> Percentile Bars for Rainfall .....	46

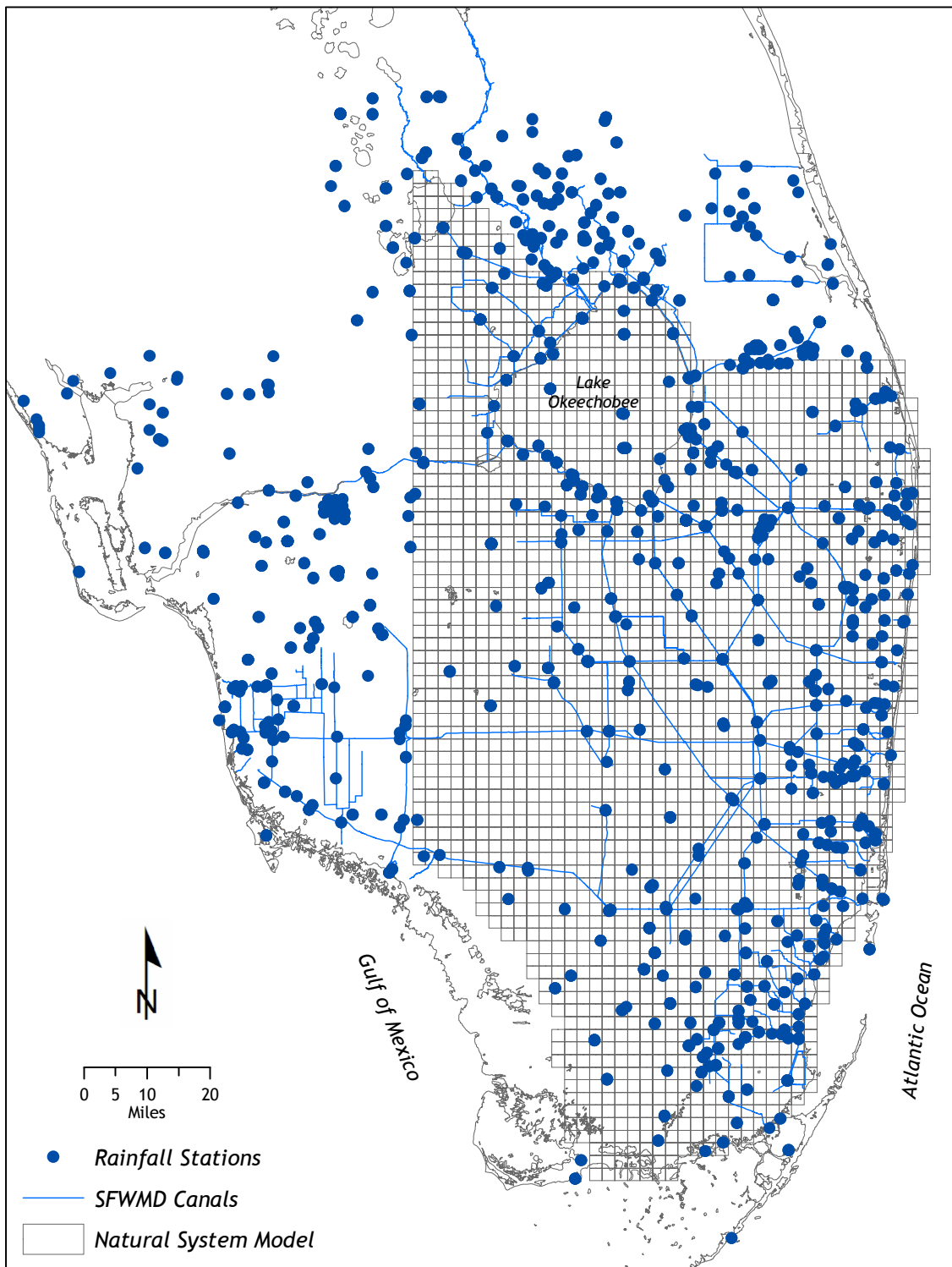
## **2.2 RAINFALL**

In all South Florida Water Management Model (SFWMM) runs, rainfall is assumed to have the same temporal and spatial distribution as that which occurred historically over the period of simulation. Since rainfall is the main driving force in the hydrology of South Florida, it serves as a good control variable for evaluating alternative ways of managing the system as a whole. For the distributed mesh portion of the model, a daily time series of rainfall depths for each grid cell is used. For Lake Okeechobee and other lumped hydrologic systems, a single daily time series of rainfall depths is input and assumed to apply over the spatial extent of the basin. The general procedure for the development of the rainfall data set in the SFWMM can be described as follows: data collection and associated quality assurance/quality control (QA/QC) or screening of rainfall station data; and transformation of rainfall point data into grid-based data.

### **2.2.1 Quality Assurance/Quality Control of Rainfall Data**

Rainfall data was collected with the goal of generating a 2-mile x 2-mile “super grid” covering nearly the entire South Florida Water Management District (SFWMD or District) for the 1914 to 2000 period of record. The spatial extent of the super grid was determined to be larger than that of the computational grid for the SFWMM in order to allow for determination of rainfall in the Natural System Model (NSM) as well as to provide rainfall information for the lumped portions of the SFWMM. The primary reason for creating a rainfall data file with a greater period of record than required by the modeling period of simulation (1965 to 2000) was to support identification of monthly and annual data trends.

Because of data availability issues, the rainfall data for the period from 1914 to 1998 were processed separately from the period of 1999 to 2000; however, the exact same procedure was used for both time periods. For the period from 1914 to 1998, there were 860 rainfall stations covering 11 counties (Broward, Highlands, Martin, Palm Beach, Collier, Glades, Monroe, Miami-Dade, Hendry, St. Lucie and Okeechobee). For the period 1999-2000, rainfall data at 964 stations covering the same counties were available. Figure 2.2.1.1 identifies the location of rainfall stations used in the creation of the SFWMM data set.



**Figure 2.2.1.1** Location of Rainfall Stations

QA/QC of rainfall station data sets was carried out in five phases, with a number of methodical steps to complete each phase. The five phases were as follows:

- I. Review and classification of daily data having extreme values.
- II. Testing and elimination of some extreme daily values.
- III. Screening of data with zero monthly rainfall.
- IV. Screening of rainfall data having extreme low annual values and high monthly values.
- V. Data screening through visualization.

The first two phases were designed to identify and remove daily values that were highly questionable according to a prescribed classification scheme, while the third and fourth phases were designed to identify and remove data associated with stations that were not consistent with monthly and annual trends. The last phase provides final QA/QC through data visualization. Appendix P presents a memorandum describing, in detail, the phases and steps used. Short descriptions of the QA/QC phases are provided in the following sections. It is important to note that during these phases, screening criteria were developed from both the raw rainfall station data and from analysis of the gridded representation of the data. The methodology for the development of the gridded data will be discussed in Section 2.2.2.

### **Phase I: Identification and Classification of Extreme Daily Rainfall Values**

In the first pass, daily rainfall values greater than 16 inches were flagged as questionable. Additionally, daily rainfall values less than 16 inches but higher than 5.5 inches in Miami-Dade, Broward and Palm Beach counties, and 5 inches in the other counties of the SFWMD area were flagged as questionable. The lower threshold values for questionable data represent approximately the 99.9 percentile in each respective county. For each day when at least one questionable data point was identified, values from the nearest six stations were extracted into a data set. For each of the resulting data sets, a classification scheme (having seven classes based on distance and value difference) was used to automatically accept or mark values for further review. After automatic acceptance of two of the classes, and marking the other five classes as questionable, the rainfall data set was recreated and reviewed using grid summaries and viewing programs.

### **Phase II: Examination of Extreme Daily Rainfall Data**

During this phase, the values identified as questionable in Phase I, were further analyzed for either acceptance or rejection. Using the nearest six stations, a manual examination of the questionable values was conducted which included consideration for: distance, direction, difference in values, number of neighbors with high values, time of year, frequency of re-occurrence in the period of record and known tropical storm events.

### **Phase III: Examination of Daily Data Corresponding to Zero Monthly Rainfall**

In this phase, efforts were made to identify and verify rainfall data for calendar months with zero rainfall. The objective was to reject or accept such data based on prescribed criteria. Part of this process was automated and part was performed manually. For each county, calendar months with zero rainfall data are extracted into a file and the average rainfall was calculated (excluding the

site under investigation) and compared to the questionable site. A monthly value of zero during dry seasons was not considered unreasonable, however zero monthly rainfall values during the wet season where nearby stations averaged  $\geq 5$  inches, were considered highly suspect. Considerations for acceptance or rejection of data included: the nearby averages, historical monthly average tables which included surrounding areas, the repetition of zero values from other sites for the same month, seasonality, the number of consecutive zero values at a given site, and whether or not the nearby site average was below the long-term monthly average. A final evaluation was made for stations with zero rainfall for three or more consecutive months by examining the quality of the daily rainfall.

#### **Phase IV: Examination of Annual Rainfall below 30 Inches and Monthly Rainfall above 20 Inches**

Visual examination of the data set showed annual rainfall was below 30 inches in some areas. Similarly, the monthly rainfall was greater than 20 inches in some areas. The examination of such data was carried out in three steps: investigation of the corresponding data, comparison with rainfall local statistics, and a visual inspection of annual snapshots extracted from the revised rainfall data set.

The investigation of the corresponding data consisted of a visual review of the daily data for the records that did not meet the criteria. About 6 percent of cases that had annual rainfall below 30 inches, 22 years of daily data were found to be of poor quality (a combination of unrealistically low and missing values) and were consequently removed. Of the cases that had a monthly rainfall that was greater than 20 inches, only month of rainfall was rejected where high rainfall was reported in an area with an average rainfall of 0.65 inches; the rest of the cases were accepted.

For the cases that had annual rainfall below 30 inches and had a maximum of two months of missing data, the following statistics were generated: the average, the standard deviation, the annual rainfall excluding the missing months, and the annual rainfall after counting for the missing month {(using the following approximation: Adjusted value = [(value)(12) / [(12 – number of missing months)]}. If the number of stations used to compute the statistics was two or less, discretion (based on a visual evaluation) was used to either reject or accept the daily data set for the year. In cases where the number of stations used to compute the statistics is more than two, the daily data set for a given year was rejected if the associated adjusted value was as follows:

1. Below 20 inches; or
2. Less than 1/2 of the average rainfall (for the given county and given year based on all locations except the one of interest); or
3. Less than (AVG-2.5)(STD) where STD is the standard deviation of annual rainfall within that county and that year.

Of the 98 cases identified, 53 daily data sets were rejected.

## **Phase V: Final QA/QC through Data Visualization**

During Phase V, a visual examination of daily, monthly, and annual snapshots of the rainfall data set was performed. Some areas of very low rainfall still existed. Associated stations were identified and a visual inspection of the daily values was performed. At some stations, daily data were of poor quality as indicated by an overwhelmingly large number of missing data for a given year. As a result of the visual evaluation, six records were rejected for at least one year, one record was rejected for two years, and three stations were dropped for the entire period of record.

### **2.2.2 Transformation to Grid-Based Data Set**

Once the rainfall data QA/QC was completed, a Triangular Irregular Network (TIN) approximation method was performed to assign a representative rainfall depth for each day and grid cell. This was necessary because rainfall gauging stations do not normally coincide with the centroid of the grid cells and most grid cells do not contain rainfall gauging stations.

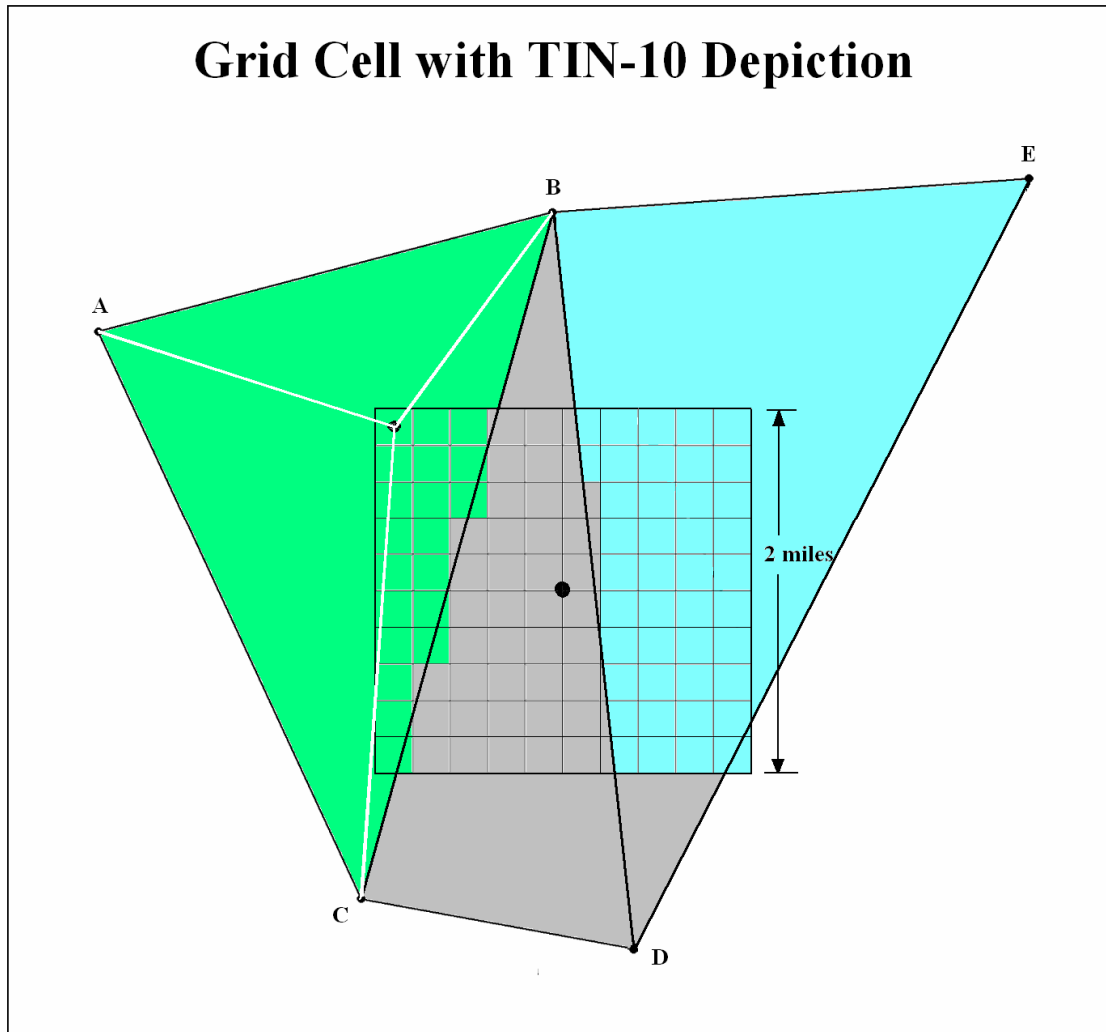
The normal TIN approximation involves using the centroid of the grid cell as a reference point for determining which three rainfall stations are used for estimating the daily rainfall value. If rainfall stations are fairly sparse, model grid cells are small, or rain events are spatially large, this would be a suitable application. However, in South Florida, the rainfall stations are not sparsely located, the model grid cells are large (4 square miles each), and heavy rainfall events can be localized. Therefore, a variation of the normal TIN approximation method was developed for this application.

The new method involved dividing each model grid cell into 100 sub-cells. Because each cell was equally divided horizontally and vertically by 10, the methodology is referred to as TIN-10. The sub-cells were over-laid by a triangular pattern of rainfall stations (with stations at each apex as shown in Figure 2.2.2.1). For the sub-cells contained within a single triangle, a daily rainfall value was calculated based on the rainfall stations at each apex. The calculated values were the weighted (based on distance from each station to each sub-cell centroid) average of the three nearest stations. Once the daily rainfall for each sub-cell was determined, the values were averaged to compute the grid cell daily rainfall value used by the model.

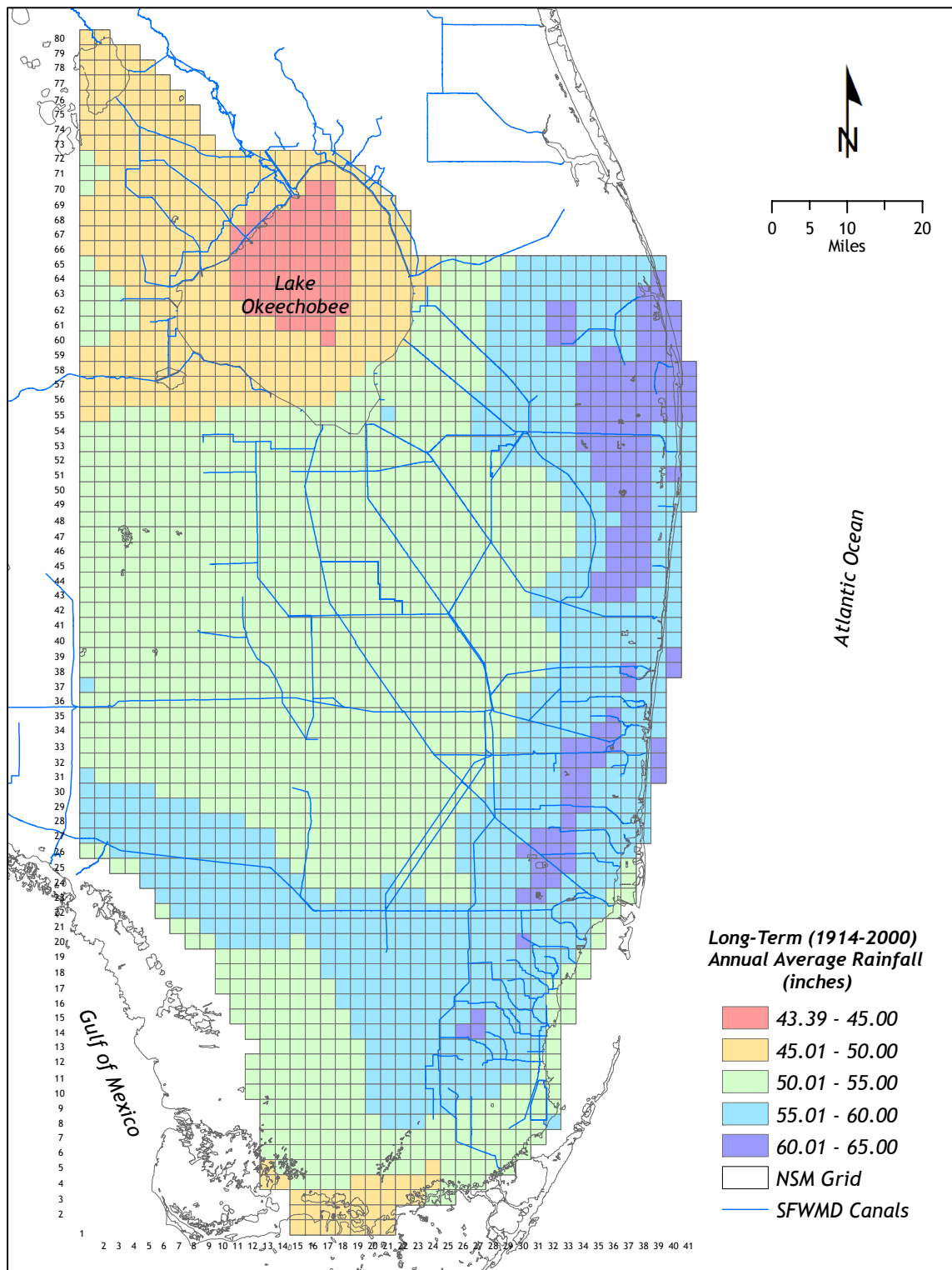
From Figure 2.2.2.1, the normal TIN approximation method would apply the rainfall at stations B, C, and D to the centroid of the grid cell even though only 38 percent of the sub-cells fell within the triangle. Consequently, the influence of two other rainfall stations would not be considered for the remaining 42 percent of sub-cells. For the TIN-10 method, the influences of the other two stations would be included in the approximation.

A comparison between the two methods revealed only small differences in annual averages with the TIN-10 method being slightly lower. The monthly average differences were generally less than 0.2 inches with the TIN-10 method having consistently lower maxima. The differences between the two methods were more evident during the wet season months. The TIN-10 method tends to decrease the dominance of any one station thus minimizing the effect of a localized rain event on a grid cell.

Average annual results of the generation of the rainfall data set by the process for data collection, QA/QC and transformation to grid are provided in Figure 2.2.2.2. The seasonal variability of the end product is shown in Figure 2.2.2.3.

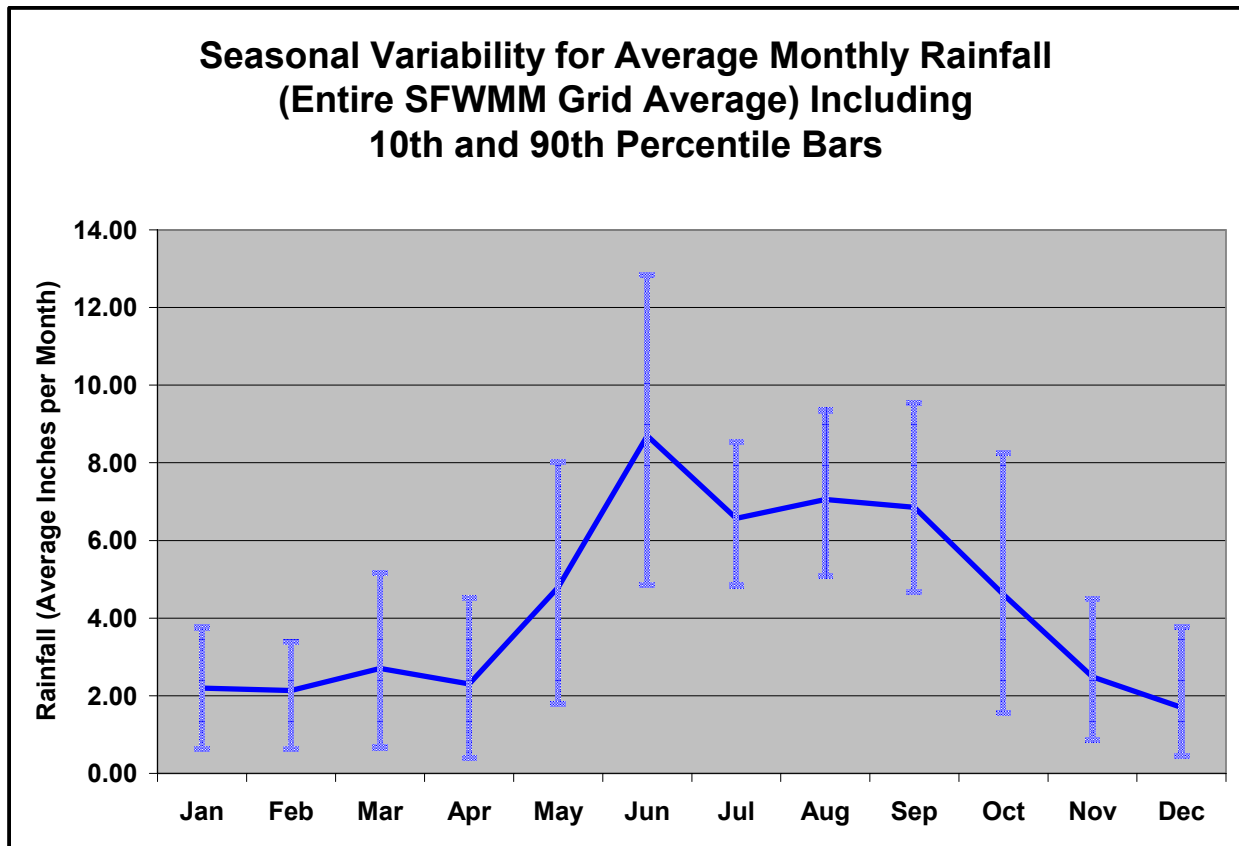


**Figure 2.2.2.1** Example of TIN-10 Estimation for Model Grid Cell



**Figure 2.2.2.2** Grid Values of Annual Average Rainfall





**Figure 2.2.2.3** Monthly Mean with 10<sup>th</sup> and 90<sup>th</sup> Percentile Bars for Rainfall